FINAL

An Experiment Using Life-size HDTV

Gloria Mark[1] and Paul DeFlorio[2]

University of California, Irvine[1] and Jet Propulsion Lab, California Institute of Technology[2]

gmark@ics.uci.edu, deflorio@mail1.jpl.nasa.gov

Introduction

NASA is interested to use technology to virtually collocate extreme collaboration teams across distance in order to share expertise. We need to explore this problem to understand how extreme collaboration[1] teams might collaborate together when teams are geographically separated. We report on an exploratory study using life-size HDTV to create a "window" across meeting rooms to connect project teams designing the same space mission.

We asked the question: what impact does high telepresence (using life-size HDTV) in virtual collocation have on team performance, when the team is distributed? We performed an exploratory study to investigate this problem using Team X[2] to see whether a virtual collocation environment would enable them to interact and conduct their work as though they were physically collocated. We split Team X up into two conference rooms.

There are several aspects of extreme collaboration teams that are important to communicate across distance. Most importantly, we felt it was essential to show activity, since monitoring activity was an important source of information.

We wanted to explore the consequences of using very high-end technologies for telepresence to overcome limitations of conventional systems. Traditional point-to-point commercial videoconferencing solutions are based on ISDN networking. A Basic Rate ISDN connection supports a data rate of 128kbps.

---

[1] See Mark (submitted paper, 2001). Extreme collaboration is defined as a project team that works together synchronously in the same electronic meeting room using computer technologies to support their information flow. Results, similar to "warroom" studies, suggest that productivity is greatly enhanced.

[2] Again, see Mark (submitted paper, 2001). Team X is one such extreme collaboration team that works at the Jet Propulsion Lab.

Traditionally, using CIF (352x288) or QCIF (176x144) images, this will support an equivalent frame rate of roughly 5-10 frames per second (fps). We refer to this as "ISDN video". Standard television (or baseband video) signals provide a much higher resolution (around 440 x 480 in the American NTSC standard) and a frame rate of 25-30 fps. However, both of these suffer from limited field of view; without introducing distortion, standard video camera have a field of view of 48.8 degrees. High-Definition Television, or HDTV, is a video standard that provides much higher resolution and a wider aspect ratio, so that the field of view is extended to 87 degrees. Because of its high frame rate, high resolution and wider field of view, HDTV would seem to provide a better basis for effective telepresence. We also chose HDTV that was life-size, wall-size to create a sense of having a "window" into another meeting room.

Experimental set-up

Static video cameras were positioned at the front of each conference room, so that each room's respective camera-based coordinate systems would be mirrored by the other (figure 1). Both Room A and B cameras were located at the left-right center of the distant room projection. The camera lenses were positioned parallel to, and about 50" above the floor, to mimic the point-of-view of a seated person. The goal of this arrangement was to support consistent spatial or gaze geometry between the two rooms. In order to maintain a human scale to the projected images, the closest people were seated a minimum of 6 feet away from the camera, while people on the close periphery (which shows more aberration caused by lens curvature) were seated at least 8 feet away. The closer objects (and people) are to the camera, the more apparent geometry mismatches become. For example, if you are standing 7 meters from the camera and point at a person displayed on the screen showing the distant room, the remote person feels that you are pointing at them, even as they move about the room. If you are 2 meters from the camera, the remote person can readily tell that you are not looking or pointing at them if they are not near the room center line.

A wide angle lens with a focal length of 5 mm was used, providing a Field-of-view (FOV) of ~56 degrees vertical by ~87 degrees horizontal. The wide FOV was needed to capture the maximum amount of floor space (people) in each room. The short focal length lens also served to keep most of the room in good static focus. The minimum object distance for viewing was about 0.5 M., which could allow for the interactive high resolution display of material objects.

In Room A (the Team X meeting room), a Panasonic high definition camcorder produced a 1920 x 1080I digital data output (60 interlaced frames a second). The camera in the second conference room (Room B) produced a 1280 x 720P digital image at 60 full frames per second. Both displays used an image width to height ratio of 16:9. The size of both projections was the same. The 16:9

image was 128 inches wide by 72 inches high with the image beginning 2 feet above the floor and continuing to within 12 inches of the ceiling. This is significantly wider than the typical image display ratio of 4:3, and conveys a larger, more immersive image of the distant room without distortion. The resolution of both projectors was 1365 x 1024 pixels. The brightness of the projectors and the sensitivity of the cameras allowed the rooms to be lit to levels well within acceptable work place standards.

Telephones were provided in each room, with telephone numbers, to support inter-site sidebars.

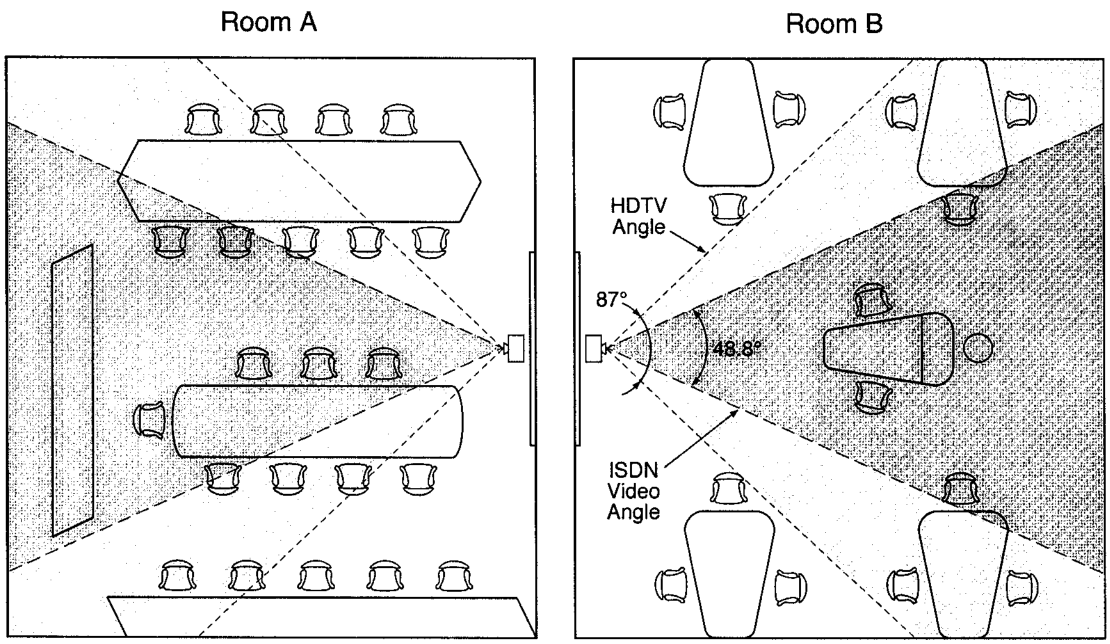

Figure 1. Top view of HDTV setup. A normal "ISDN video" FOV is shown to compare.

Experimental conditions: Lag vs. synchronicity

On day 1, a single customized binaural (stereo) microphone, placed around the video camera body, sent audio directly into the other room where it was reproduced via an amplifier and two tall floor-to-ceiling speakers. The video was compressed using MPEG2 and sent through a local Gigabit fiber network via Ethernet IN a 19.4 Megabit/second data stream of IP encapsulated MPEG packets and auxillary management data. There was a .8 second lag of video to audio.

On day 2 both the audio and video were digitized. The one-way audio transmission sequence was: microphone input, analog to digital, MPEG encode, ip encapsulate, transmit, receive, and the reverse to stereo amplifier and distant speakers. The video was transmitted as in Day 1. This resulted in a synchronous audio and video signal, but the audio signal had acoustic feedback, which was corrected halfway through the meeting.

Experiment

The experiment was conducted over two Team X sessions on Tuesday and Thursday of one week, three hours each session. The team was hired by a customer to determine the feasibility of a space shuttle mission, so it was an actual working session. In Room A were eight team members and the customers. In Room B were seven members and the team leader. Both authors observed the team, and activity was coded. Video-taping was not permitted on Day 1 by the customer, but was done on Day 2. Questionnaires were distributed to the team members after Day 2, and a one-hour group discussion was held with Team X.

Results

Overview of results

The team leader sat in Room B (fig. 3) and is shown in the HDTV image (fig. 2). In the authors' view, the life-size HDTV video produced an extremely compelling image in terms of high resolution, and showing breadth of the remote room. Some representative comments of the team support this:

> The HDTV projection was very good, as I stated during the session. The details gave the HDTV projection almost 3-D like view.
>
> The quality of the picture on both days was outstandingly clear and crisp.
>
> The video was great, you could make out facial features of people at the back of the room. The only problems were that the camera was in the middle of the viewing screen and sometimes blocked the images.

Fig. 2. Room A in the HDTV experiment. The HDTV image appears small in the photo due to the wide-angle camera lens used for the photo. The team leader in the image in fact appeared life-size.

Fig. 3. Room B in the HDTV experiment. The team leader is sitting in the foreground.

Perhaps one of the indications that video created a sense of telepresence was that team X members turned in their seats to face the video screen when speaking to remote team members, using social conventions as in face-to-face interaction.

The audio created a spatial audio effect, i.e. it was easy to discern one's seating location in the remote room when a remote person spoke. However, when the team leader had his back to the camera, it was hard to hear him. Telephones were available (with phone numbers) to support inter-room sidebars, but were not used.

Differences in Day 1 (lag) vs. Day 2 (synchronicity)

	Day 1	**Day 2**	**Face-face meeting**
1=Extremely poor, 10=extremely high			
Quality of audio	M=6.0 (sd=2.2)	M=5.3 (sd=2.3)	
Quality of video	M=7.8 (sd=1.0)	M=8.1 (sd=1.0)	
Presence	M=6.3 (sd=3.2)	M=6.1 (sd=3.0)	
Satisfaction	M=5.7 (sd=2.1)*	M=5.9 (sd=2.1)	M=9.4 (sd=.9)*
Scale for following: 1=not a problem, 10=extreme problem			
Lag on Day 1	M=4.8 (sd=2.6)		
Audio fb on Day 2		M=7.2 (sd=1.9)	
Sidebars	M=8.0 (sd=1.9)	M=7.8 (sd=1.9)	
*p < .01			

Table 1. Questionnaire results for HDTV experiment.

Table 1 shows questionnaire results from some of the questions. The quality of the audio was judged by the team to be moderate on both days, and the HDTV video was judged higher, though not as high as we expected. The engineers were asked how “present“ they felt the team members in the other room were, and judged this to be moderately high. Audio feedback was judged to be more of a problem on Day 2 than the .8 second lag on Day 1. The team members were asked which technology experience they were more satisfied with, Day 1 or Day 2. Two members preferred Day 2, one member felt they were about the same, and the rest of the team preferred Day 1. Not unexpected, audio feedback was less tolerable than video that was out of synch with the audio.

Team members were far less satisfied with the distributed meetings than with a typical face-to-face meeting. There was no difference in satisfaction for the two experiment days, so results were combined and a paired-sample T-test showed the difference compared to face-to-face was highly significant $t(11)=-4.4$, $p<.004$.

Communication and sidebar conversations

As described earlier, the heart of the Team X experience is the sidebar conversations. How well did the HDTV support sidebar conversations when the team was distributed? First of all, it is important to keep in mind that interdependencies did exist among the members *between* rooms, and not just among members within rooms. The telephones were not used for sidebars. We first looked at what people reported they did. In the questionnaire, team members were asked what they did when they needed to engage in a sidebar discussion. Their choices were: I did nothing, I walked over to a person in the same room to have a sidebar conversation, I walked over to a person in the other room to have a sidebar conversation, I spoke to the person through the technology connection, or other action. Figure 4 shows the distribution of responses for Day 1 and Day 2. On Day 1, members claimed that on the average, 12% of the time they did nothing when they needed to have a sidebar, and perhaps through learning, this dropped to about 5% on Day 2. Both days, about 15% of the time the members reported that they walked into the other conference room. This behavior is further evidence that the technology did not support sidebars, as they had a real customer and did what they felt they needed to do to continue their work. Sidebars could be held using the HDTV and audio connection, and this was reported to have been done only about 12-15% of the time when sidebars were needed.

Sidebars were coded from the videotape of the Day 2 session, which clearly showed both rooms. In fact, compared to the 98 average number of sidebars coded during face-to-face meetings, only 74 sidebars occurred during the HDTV meeting. Fifty-six sidebars occurred in the same room, and 18 sidebars occurred when people crossed over into the other room to meet face-to-face.

The main problem that members reported was that the video/audio did not support sidebar conversations (Table 1). Some comments were:

Instead of being able to walk over to the person, tap them on the shoulder, and ask them a question, you had to call out their name over the audio system in *order to get their attention. Hopefully the person heard you, otherwise you had to try again.*

When you are all in the same room, everyone can see each other, hold discussions, and the leader can monitor what's going on very effectively. When you are located in separate rooms, side discussions become an obstacle that you have to work out.

I had to 'cheat' on several occasions and walk next door to talk to my counterparts, and I noticed everyone else was doing that too.

Two things need to be added to make it reasonable. A way of switching off the microphones at the workstations so private sidebars don't interfere with the rest of the team, and an easy means for a person at one workstations to communicate easily with someone at another workstation at a remote location...

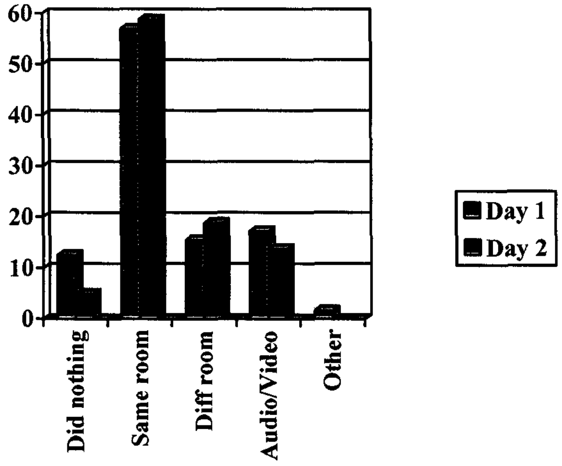

Figure 4. Estimates of sidebar activity during HDTV experiment.

The single channel as information conduit

The team leader became a single primary channel of information for the group, as opposed to normal meetings, where he is wandering around the room, checking that people were publishing and subscribing as needed. The only information that he had as an overview of the team progress was the spreadsheet on the public display. He could see who was talking to whom, but he did not have detailed information, e.g. what was on people's screens, and sidebar conversations.

If the HDTV/audio did not suitably support the team in terms of its sidebar conversations, what function did it have? The HDTV/audio appeared to have two primary functions for the team. First, it supported public conversation. As a single, directed channel, the HDTV/audio was very effective. Some sidebar conversations did occur through the HDTV/audio channel, but they became in fact, public conversations due to the nature of the audio channel.

Second, it functioned as an awareness mechanism for activity in the other room. The HDTV FOV showed all the action in the remote room. Perhaps one reason that the lag of video to audio was not judged very disturbing in Day 1, was that the HDTV image functioned primarily as visual awareness, rather than as a communication channel for the entire team. Team members saw who was speaking with whom through the HDTV image, and then walked into the other room (18 times) to join the conversation. However, activity through the HDTV image is not as salient as activity in the actual conference room, despite the best efforts to create high quality life-size video.

Discussion and implications

A virtual collocation environment as a window between conference rooms is a start, but is not enough. Neither telephones nor the single channel sufficiently supported sidebar interaction of members in different rooms. Interaction between sites requires far more effort than interaction within sites. The large ratio of within-room to between-room sidebar conversations support this, as well as the questionnaire responses. The physical collocation of team X is needed to change the raw numbers into knowledge that can be further processed. By missing out on between-room sidebars, the necessary human collaborative processing may not have been done during the experiment. This is a problem to be aware of in a virtual collocation setting. Within-room interaction will always be easier, and distributed work may suffer by losing out on between-room interaction.

It was our aim by using wall-size, life-size HDTV to overcome many of the problems found in interaction with regular video-conferencing systems. First, the HDTV conveyed a far wider angle of the remote room with less distortion, than normal video-conferencing. Members could move within a wide range in the room and still be captured by the video. Second, the resolution was a much higher quality than normal video. We did not observe any exaggerated gestures or movements to convey expression through the HDTV image, as Heath and Luff (1991) found with normal video. As mentioned, one engineer said that he saw the facial expression of a remote team member seated in the back of the remote room.

In the HDTV experiment, there were fewer sidebars overall (as a result of few between-room sidebars), which could have several implications for virtual collocation. First, the amount of automated information processing could be higher. Second, less oral processing of information could impact the quality of the results, as less options are explored, less assumptions are questioned, and less errors are caught and corrected. Third, members between rooms must rely on the single channel (team leader) for information on progress. There is the higher chance that the two meeting rooms can fall out of synch due to the limited information flow between rooms. The burden falls more on the team leader to keep both teams assessed of progress, changes in the design, and errors.

It is a challenge in a virtual collocation environment to support not only intentional sidebars across sites but also spontaneous sidebars. The HDTV video may show who is speaking, but there is so much audio information even in the same room that it is hard to discern content across distance. Even though the video shows who is talking with whom, the act of making the connection across distance loses the spontaneity that Team X has in a face-to-face environment.

Stress has been reported as a problem by many of the team members. There is a large amount of information processing occurring during a session. By connecting teams across distance, e.g. through an HDTV/audio channel, it increases the amount of information processing even more, when it opens up a window into another room. It is more difficult to attend to the information on a screen, compared to physical activity in the same room.

Design recommendations

We plan to continue testing life-size HDTV with distant institutions, such as the European Space Agency, Centre Nationale d'Etudes Spatiale, and other NASA sites. There are a number of opportunities for experimenting with new technologies to support virtual collocation for extreme collaboration:

- Wireless headsets could support sidebar conversations across distance
- A simple drag-and-drop computer interface showing a seating map could help users seamlessly establish connections between remote people.
- The illusion of spatial audio could be enhanced by relocating speakers behind a flexible screen; the sound would appear to be coming from the other room.
- Earphones to selectively mute the ambient noise would relieve stress, although it reduces the opportunities to monitor conversation.
- Camera presence could be minimized, e.g. by fabricating a stand to orient the camera body upward, and then attaching a "periscope" to raise and correctly orient the camera lens at its end.
- Transmit and receive buffer size tuning is a trade-off between delay (with bigger buffer size) vs. robustness (the impact of IP packet loss on the decode/display side). This can be a real-time variable of the system.

Technology design for virtual collocation needs to consider finer-grained support for interdependencies among group members, in terms of information exchange, and not just a broad channel. With extreme collaboration, the information flow changes rapidly in different permutations, and a communication medium must support this.

Acknowledgements

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration, and by the University of California at Irvine

References

Heath, C. and Luff, P. (1991). Disembodied conduct: Communication through video in a multi-media office environment. Proc. of CHI'91, pp. 99-103.